بسم الله الرحمن الرحیم

شرکت مهندسی نرم افزاری هلو

گزارش مربوط به تولید مجموعه های داده­ای مربوط به Action Recognition

کاری از امیرعلی نسیمی

* **مقدمه**

با توجه به اهمیت پردازش فایل های ویدئویی مربوط به شناسایی Command، ایده­ی شناسایی Action در فایل های ویدئویی مورد بررسی قرار گرفت. با توجه به این مسئله، یکسری از مشهورترین مجموعه­های داده­ای در قالب فایل ویدئویی در این گزارش معرفی شده است.

* **[UCF101 (UCF101 Human Actions dataset)](https://www.crcv.ucf.edu/data/UCF101.php)**

مجموعه داده UCF101 توسعه ای از UCF50 بوده و شامل 13320 کلیپ ویدیویی است که در 101 دسته طبقه بندی شده­اند. این 101 دسته را می توان به 5 نوع (حرکت بدن، تعامل انسان و انسان، تعامل انسان و شی، نواختن آلات موسیقی و ورزش) طبقه بندی کرد. طول زمانی کل کلیپ های ویدیویی مذکور به بیش از 27 ساعت ضبط شده است. تمامی ویدیوها از یوتیوب جمع آوری شده و دارای نرخ فریم ثابت 25 فریم در ثانیه با رزولوشن 320 × 240 هستند.

* [**Kinetics (Kinetics Human Action Video Dataset)**](https://www.deepmind.com/open-source/kinetics)

مجموعه داده­ای Kinetics یک مجموعه داده با مقیاس بزرگ و با کیفیت بالا برای تشخیص اقدامات انسانی در انواع ویدیو می­باشد. مجموعه داده­ای مذکور شامل حدود 500000 کلیپ ویدیویی است که 600 کلاس عمل انسانی را با حداقل 600 کلیپ ویدیویی برای هر کلاس اکشن پوشش می دهد. هر کلیپ ویدیویی حدود 10 ثانیه بوده و با یک کلاس اکشن برچسب گذاری شده است. فیلم ها از یوتیوب جمع آوری شده­اند.

* [**HMDB51**](https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/)

مجموعه داده HMDB51 مجموعه بزرگی از ویدیوهای واقعی از منابع مختلف، از جمله فیلم­ها و ویدیوهای موجود در اینترنت است. مجموعه داده مذکور از 6766 کلیپ ویدیویی بصورت 51 دسته اکشن انسانی (مانند "پرش"، "بوسه" و "خنده") تشکیل شده است که هر دسته شامل حداقل 101 کلیپ می­باشد. جهت ارزیابی نسبت 70 به 30 برای هر کلاس رعایت شده است.

* [**ActivityNet**](http://activity-net.org/)

مجموعه داده ActivityNet شامل 200 نوع فعالیت مختلف و در مجموع 849 ساعت ویدیو جمع آوری شده از YouTube است. ActivityNet بزرگترین معیار برای تشخیص فعالیت­های زمانی تا به امروز - از نظر تعداد دسته های فعالیت و تعداد ویدیوها بوده که این کار را به ویژه چالش برانگیز می کند. نسخه 1.3 مجموعه داده در مجموع شامل 19994 ویدیو بدون ترمیم شده است و به سه زیرمجموعه مجزا، آموزش، اعتبارسنجی و آزمایش با نسبت 2:1:1 تقسیم شده است. به طور متوسط، هر دسته فعالیت دارای 137 ویدیوی بدون برش است. هر ویدیو به طور متوسط دارای 1.41 فعالیت است که با مرزهای زمانی حاشیه نویسی شده است. حاشیه‌نویسی‌های واقعی ویدیوهای آزمایشی عمومی نیستند.

* [**Charades**](http://vuchallenge.org/charades.html)

مجموعه داده Charades از 9848 ویدیو از فعالیت‌های روزانه در داخل خانه با تعداد فریم­های 30 در 1 ثانیه تشکیل شده است که شامل 46 کلاس شی در 15 نوع صحنه داخلی و حاوی برچسب­هایی از 30 عمل است که منجر به 157 کلاس اکشن می‌شود. هر ویدیو در این مجموعه داده با توضیحات متن آزاد متعدد، برچسب‌های کنش، فواصل عمل و کلاس‌های اشیاء متقابل حاشیه‌نویسی می‌شود. به 267 کاربر مختلف جمله ای ارائه شد که شامل اشیاء و اعمال از یک واژگان ثابت است و آنها فیلمی را با اجرای این جمله ضبط کردند. در مجموع، مجموعه داده شامل 66500 حاشیه نویسی زمانی برای 157 کلاس عمل، 41104 برچسب برای 46 کلاس شی و 27847 شرح متنی ویدیوها است. در تقسیم استاندارد 7986 فیلم آموزشی و 1863 ویدیوی تأیید وجود دارد.

* [**MPII (MPII Human Pose)**](http://human-pose.mpi-inf.mpg.de/)

مجموعه داده‌های MPII Human Pose برای تخمین ژست تک نفره از حدود 25 هزار تصویر متشکل از 15هزار نمونه آموزشی، 3 هزار نمونه‌ی Validation و 7 هزار نمونه‌ی آزمایشی می­باشد (دسترسی به برچسب ها محدود می­باشد). تصاویر مربوطه از ویدئوهای یوتیوب گرفته شده است که 410 فعالیت مختلف انسان را پوشش داده و ژست ها به صورت دستی با حداکثر 16 مفصل بدن برچسب زده شده­اند.

* [**NTU RGB+D**](https://rose1.ntu.edu.sg/datasets/actionrecognition.asp)

NTU RGB+D یک مجموعه داده در مقیاس بزرگ برای تشخیص اقدامات انسانی است. این مجموعه شامل 56880 نمونه از 60 کلاس بوده که از 40 نفر جمع آوری شده است. این اعمال را می توان به طور کلی به سه دسته تقسیم کرد: 40 عمل روزانه (مانند نوشیدن، خوردن، خواندن)، 9 عمل مرتبط با سلامتی (مانند عطسه، تلوتلو خوردن، افتادن) و 11 عمل متقابل (مانند مشت، لگد، در آغوش گرفتن). این اقدامات تحت 17 شرایط صحنه مختلف مربوط به 17 دنباله ویدئو (به عنوان مثال، S001-S017) انجام می شود. اقدامات با استفاده از سه دوربین با نماهای تصویربرداری افقی مختلف گرفته شد. اطلاعات چند وجهی مختلفی از قبیل نقشه های عمق، موقعیت مفصل اسکلت سه بعدی، فریم های RGB و دنباله های مادون قرمز برای توصیف عملکرد ارائه شده است.

* [**KTH (KTH Action dataset)**](https://www.csc.kth.se/cvap/actions/)

تلاش ها برای ایجاد یک مجموعه داده غیر پیش پا افتاده و در دسترس عموم برای شناسایی کنش ها در مؤسسه فناوری KTH در سال 2004 آغاز شد. مجموعه داده KTH یکی از استانداردترین مجموعه داده ها است که شامل شش عمل است: راه رفتن، دویدن، راه رفتن، جعبه، موج دست، و کف زدن دست. برای در نظر گرفتن تفاوت های ظریف عملکرد، هر عمل توسط 25 فرد مختلف انجام شده و در نهایت تنظیمات به طور سیستماتیک برای هر عمل به ازای هر بازیگر تغییر می کند. تغییرات تنظیمات عبارتند از: فضای باز (s1)، فضای باز با تغییر مقیاس (s2)، فضای باز با لباس های مختلف (s3)، و فضای داخلی (s4). این تغییرات توانایی هر الگوریتم را برای شناسایی اقدامات مستقل از پس زمینه، ظاهر بازیگران و مقیاس بازیگران آزمایش می کند.

* [**Sports-1M**](https://cs.stanford.edu/people/karpathy/deepvideo/)

مجموعه داده Sports-1M شامل بیش از یک میلیون ویدیو از YouTube است. ویدیوهای موجود در مجموعه داده را می توان از طریق URL YouTube مشخص شده توسط نویسندگان به دست آورد. تقریباً 7٪ (از سال 2016) از ویدیوها توسط آپلودکنندگان YouTube از زمان گردآوری مجموعه داده حذف شده است. با این حال، هنوز بیش از یک میلیون ویدیو در مجموعه داده با 487 دسته مرتبط با ورزش با 1000 تا 3000 ویدیو در هر دسته وجود دارد. ویدیوها به طور خودکار با 487 کلاس ورزشی با استفاده از YouTube Topics API با تجزیه و تحلیل فراداده متنی مرتبط با ویدیوها (به عنوان مثال برچسب ها، توضیحات) برچسب گذاری می شوند. تقریباً 5٪ از ویدیوها با بیش از یک کلاس حاشیه نویسی شده اند.

* **تحلیل مجموعه­های داده**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| مجموعه داده | تعداد کل فایل­ها | تعداد کل کلاس ها | حوزه | ساختار دقیق برچسب ها | نمونه | امکان دانلود | میزان وفق پذیری با پروژه |
| UCF101 | 13320 کلیپ ویدیویی | 101 | 1. حرکت بدن 2. تعامل انسان و انسان 3. تعامل انسان و شی 4. نواختن آلات موسیقی 5. ورزش | [لینک](https://www.crcv.ucf.edu/data/UCF101.php) | [لینک](https://www.crcv.ucf.edu/data/UCF101/UCF101.jpg) | [**لینک**](https://www.crcv.ucf.edu/data/UCF101/UCF101.rar) | **متوسط** |
| Kinetics-700 | 650000 کلیپ ویدئویی | 700 | تعامل انسان | بغل کردن، دست دادن و .. | [لینک](https://production-media.paperswithcode.com/datasets/kinetics700.jpeg) | [**لینک**](https://storage.googleapis.com/deepmind-media/Datasets/kinetics700_2020.tar.gz) | **متوسط** |
| Kinetics-600 | 600 | [**لینک**](https://storage.googleapis.com/deepmind-media/Datasets/kinetics600.tar.gz) |
| Kinetics-400 | 400 | [**لینک**](https://storage.googleapis.com/deepmind-media/Datasets/kinetics400.tar.gz) |
| HMBD51 | 7000 کلیپ ویدئو | 51 | 1. رفتار های چهره 2. رفتار های چهره با شی 3. رفتار های اندام 4. تعامل انسان با انسان | لینک [1](https://serre-lab.clps.brown.edu/wp-content/uploads/2012/08/HMDB_snapshot1.png) و [2](https://serre-lab.clps.brown.edu/wp-content/uploads/2012/08/HMDB_snapshot2.png) | لینک [1](https://serre-lab.clps.brown.edu/wp-content/uploads/2012/08/HMDB_snapshot1.png) و [2](https://serre-lab.clps.brown.edu/wp-content/uploads/2012/08/HMDB_snapshot2.png) | [**لینک**](http://serre-lab.clps.brown.edu/wp-content/uploads/2013/10/hmdb51_org.rar) | **متوسط – بخشی از نمونه­ها لازم نیست** |
| ActivityNet | 20000 ویدئو کلیپ | - | [لینک](https://todo.tnc.ir/secure/attachment/26495/Screenshot%202023-07-19%20152418.jpg) | [لینک](https://todo.tnc.ir/secure/attachment/26495/Screenshot%202023-07-19%20152418.jpg) | [لینک](https://todo.tnc.ir/secure/attachment/26496/Screenshot%202023-07-19%20152641.jpg) | لینک [1](https://docs.voxel51.com/user_guide/dataset_zoo/index.html)  و [2](http://ec2-52-25-205-214.us-west-2.compute.amazonaws.com/files/activity_net.v1-3.min.json) - هر چند دانلود این مجموعه بسیار مشکل می­باشد. | **مناسب** |
| Charades | 9848 ویدئو کلیپ | - | تعامل انسان | 157 برچسب | [لینک](http://vuchallenge.org/images/charades-bg.jpg) |  | **عدم امکان دانلود** |
| MPII | 25000 کلیپ ویدئویی | 410 | لینک [1](https://todo.tnc.ir/secure/attachment/26498/Screenshot%202023-07-19%20160614.jpg) و [2](https://todo.tnc.ir/secure/attachment/26499/Screenshot%202023-07-19%20160627.jpg) و [3](https://todo.tnc.ir/secure/attachment/26500/Screenshot%202023-07-19%20160704.jpg) | [لینک](http://human-pose.mpi-inf.mpg.de/images/random_activities.png) | [لینک](http://human-pose.mpi-inf.mpg.de/images/random_activities.png) | لینک [1](https://datasets.d2.mpi-inf.mpg.de/andriluka14cvpr/mpii_human_pose_v1_sequences_batch1.tar.gz) و ..  حدود 25 فایل | **مناسب** |
| NTU RGB+D | 57000 کلیپ ویدئو | 60 | **به دلیل عدم کار وبسایت مربوطه، اطلاعاتی در این خصوص وجود ندارد** | | | | |
| KTH | 2391 sequences | 6 | 1. راه رفتن 2. دویدن 3. راه رفتن 4. جعبه 5. موج دست 6. کف زدن دست | [لینک](https://www.csc.kth.se/cvap/actions/actions.gif) | [لینک](https://www.csc.kth.se/cvap/actions/actions.gif) | لینک [1](http://www.csc.kth.se/cvap/actions/walking.zip) و ...  حدود 12 | **نامناسب** |
| Sports-1M | 1,133,158 | 487 | ورزش | [لینک](https://cs.stanford.edu/people/karpathy/deepvideo/sz70h.jpg) | [لینک](https://cs.stanford.edu/people/karpathy/deepvideo/sz70h.jpg)  [لینک](https://www.youtube.com/watch?v=UDqivjS-lpI) | [**Git clone**](https://github.com/gtoderici/sports-1m-dataset/) | **نامناسب** |